

Poster III-4

Can Hybrid Alignment Enhance PSI-BLAST?

Li, Y., Lauria, M., Bundschuh, R.*

Ohio State University, Columbus, OH, USA

The most fundamental and widely used computational tool in the biomedical discovery process is sequence alignment. Although sequence alignment is a well-established technique, the need to detect weaker and weaker sequence homologies requires continuous improvements in the sensitivity of alignment algorithms. We are investigating new methods to increase the sensitivity of protein sequence alignment and we use the popular tool PSI-BLAST to put them to test. PSI-BLAST is already now often the program of choice for high sensitivity database searches. However, a number of researchers still find alternative manual procedures to be more appropriate for identifying extremely weak homologies. What puts PSI-BLAST at a disadvantage is that it is oblivious of the different propensities for insertions and deletions (i.e., gaps) in different regions of the sequences. This is a fundamental limitation of the underlying statistical theory, which cannot take into account position-specific gap costs.

Recently, a novel alignment algorithm has been proposed that offers features expected to improve the sensitivity of such iterative approaches. This so-called hybrid algorithm features a well-characterized theory of its statistics even in the presence of position-specific gap costs. Replacing the core of PSI-BLAST with the new algorithm will produce a position-specific gap cost aware version of PSI-BLAST. However, it is not clear how the new algorithm will interact with the rest of the PSI-BLAST framework, which has been heavily optimized for its original alignment core.

We test the feasibility of this approach by partially replacing the alignment core of PSI-BLAST with the hybrid algorithm. In this way we establish the following results. First, the hybrid alignment algorithm can be successfully used within PSI-BLAST with only modest changes to the original code. Therefore, we are able to effectively leverage all the efforts that went into the development of the current PSI-BLAST and the tools that build upon it. Second, in studying how to best match the algorithm to the PSI-BLAST code, we resolve the question of sequence length correction. Third, through direct comparison, and in optimizing one parameter for the hybrid algorithm within the whole framework of PSI-BLAST, namely the gap costs, we find that the new version of PSI-BLAST and the original version of PSI-BLAST are very similar in their performance.

These findings provide a basis for future exploitation of features of the hybrid algorithm that the Smith-Waterman algorithm does not provide. Most notably, they open the possibility of including position-specific gap costs in PSI-BLAST.

This work is supported by the National Science Foundation under grant DBI-0317335.